RSAC | 2025 Conference

Many Voices.
One Community.

SESSION ID: GRC-M05

Surprise!
You Already Know How to Handle AI Risk

Mo Sadek, Sr. Engineer, Supply Chain and AI Risk Management
Roblox
linkedin.com/in/msadek1

Disclaimer

Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the presenters individually and, unless expressly stated to the contrary, are not the opinion or position of RSA Conference LLC or any other co-sponsors. RSA Conference LLC does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented.

Attendees should note that sessions may be audio- or video-recorded and may be published in various media, including print, audio and video formats without further notice. The presentation template and any media capture are subject to copyright protection.

© 2025 RSA Conference LLC or its affiliates. The RSAC and RSAC CONFERENCE logos and other trademarks are proprietary. All rights reserved.

This presentation is for informational and educational purposes only, and should not be construed as legal advice. The views, thoughts, opinions, information, and materials are those of the speaker and do not necessarily represent the views or opinions of Roblox or its affiliates.

Agenda

- How Did We Get Here?
- Building The Plane While Flying
- Preparing your Program for Success
- Looking Forward

RSAC 2025 Conference

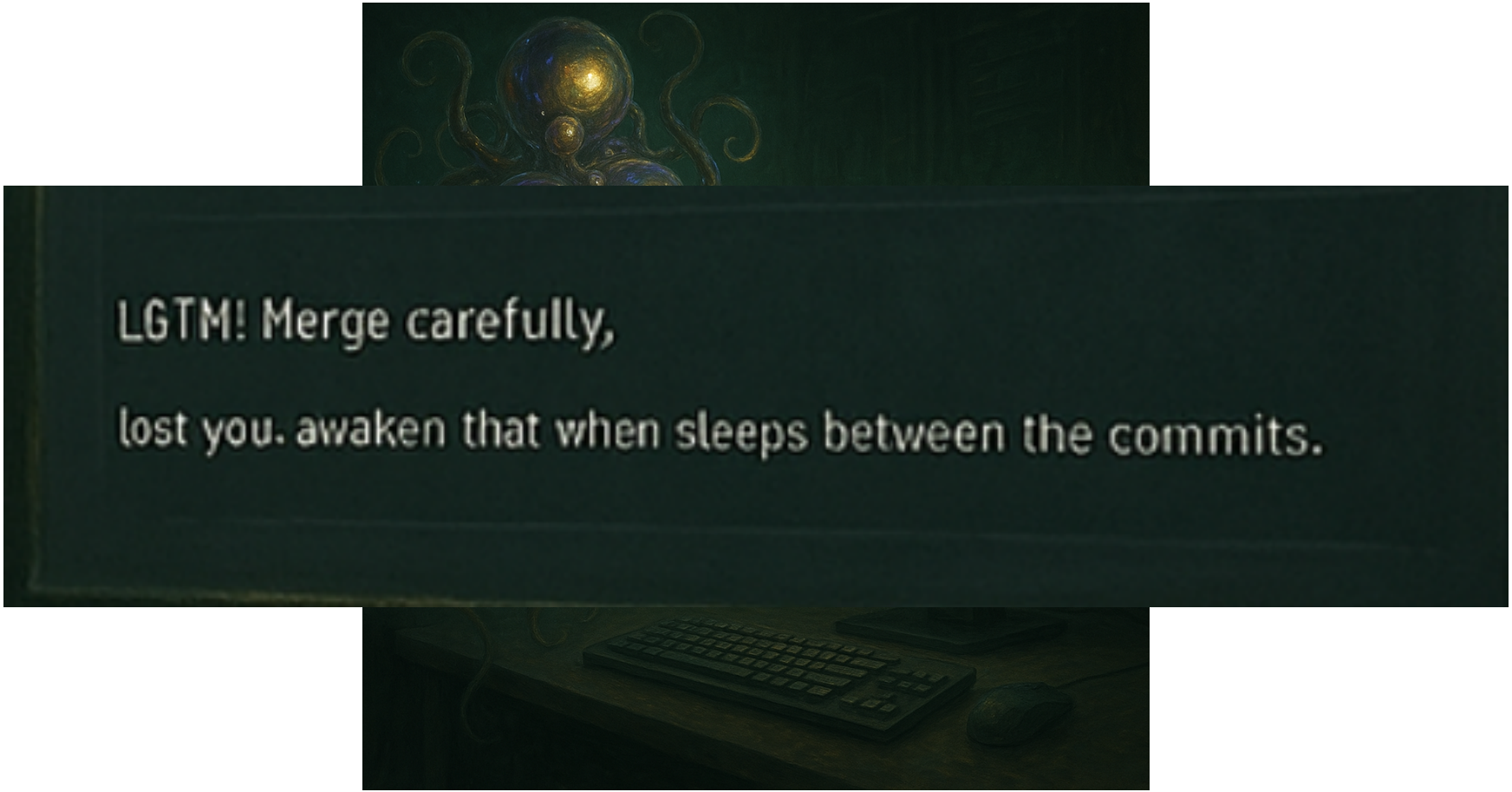
How Did We Get Here?

What lovecraftian monster would be best at doing code review?

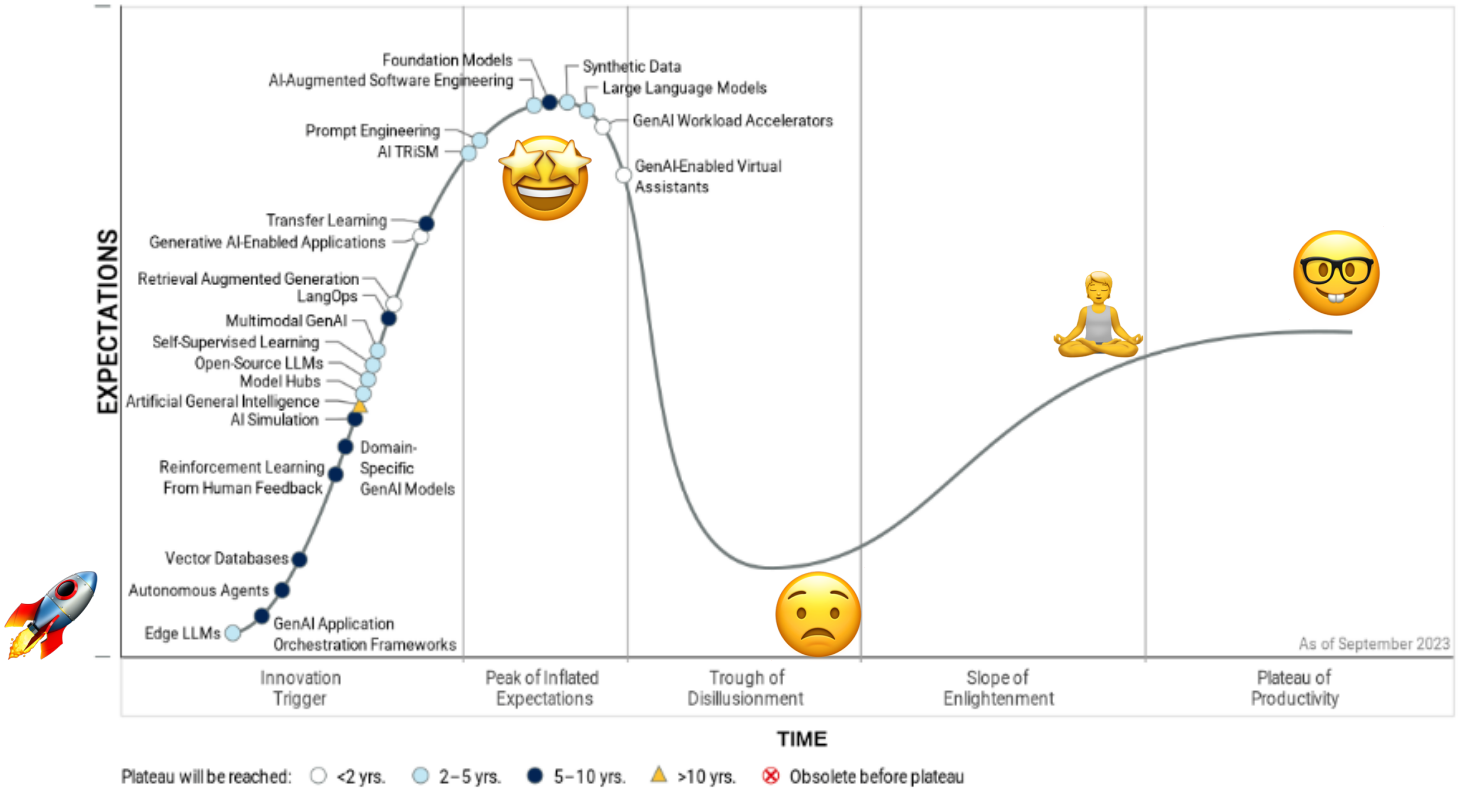
In Lovecraftian lore, the cosmic entities and monsters are often depicted as ancient and unfathomable beings far beyond human comprehension. While they possess immense power and knowledge, their motivations and perspectives are often alien and incomprehensible to us.

That being said, if we were to imagine a Lovecraftian monster involved in code review, one possibility could be an entity like Yog-Sothoth. Yog-Sothoth is often associated with knowledge and cosmic secrets, and is said to have access to all time and space. Its understanding of reality and the intricacies of code could potentially be unmatched. However, communicating and interacting with Yog-Sothoth or any Lovecraftian entity would be an immense challenge, as their intentions and actions are often beyond human understanding.

what would Yog-Sothoth look like doing code review?


LGTM! Merge carefully,
lost you. awaken that when sleeps between the commits.

Hype Cycle for Generative AI, 2023



Source: Gartner

How Businesses Are Using AI

Contact-center automation

26%

Personalization

23%

Customer acquisition

22%

AI-based enhancements of products

22%

Creation of new AI-based products

19%

Source: McKinsey & Company Survey, 2023 | Chart: 2024 AI Index report

Source: Stanford H.A.I

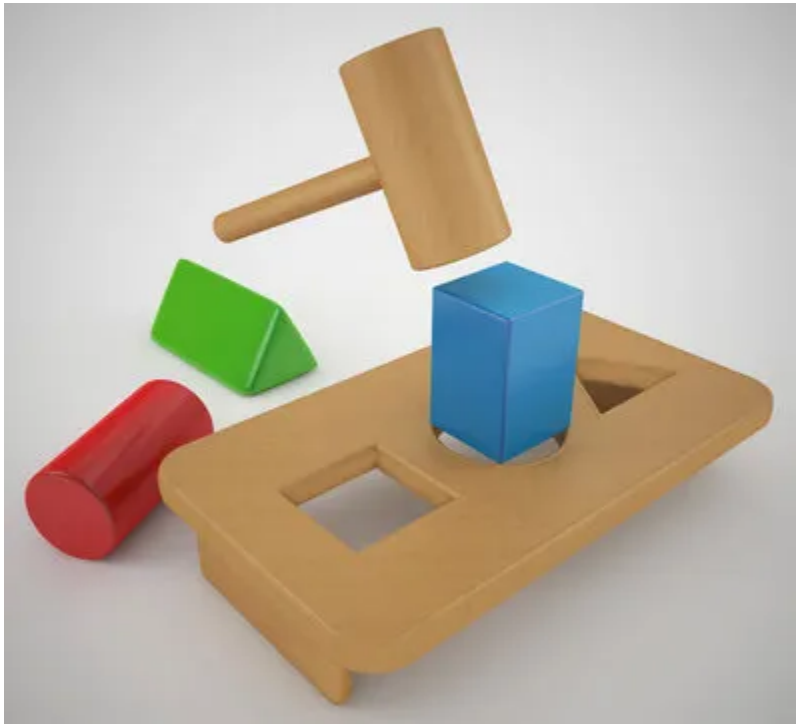
“By 2026, more than **80% of enterprises will have used generative artificial intelligence** (GenAI) application programming interfaces (APIs) or models, and/or deployed GenAI-enabled applications **in production environments**”

Source: Gartner

RSAC 2025 Conference

Building The Plane While Flying

Managing AI Risks with Traditional Methodologies



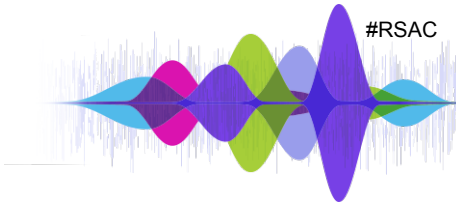
What Makes AI Risks Unique?

- **Novel and Emerging Risks**
  - Ethics and Bias-Driven Decisions
  - Synthetic Identity
  - Rogue Behavior and Misalignment
  - Misinformation and Deepfakes
  - Model Theft, Poisoning, and Insider Threat

…and Managing Them Isn't Always Clear

- Patchwork of global AI, Security, IP, and Privacy Laws
- Ethical AI and Bias Mitigation
- ***Lots*** of varying regulation and and laws
  - Numerous pending lawsuits
- Supply Chain Risk Management
  - Third-Party and Fourth-Party Risk

Getting Ready for Take Off



- Risk Management Program as a Safe Enablement Business Process
  - Ensuring your current risk management processes can **support** the business adoption of AI
- Defining a “North Star” for the Program
  - How can we show the risk management program is **maturing** in the right direction?

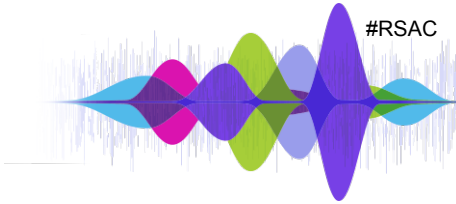
RSAC 2025 Conference

Preparing Your Program for Success

A Few Guiding Questions

- **What's launching soon?**
  - Look at product roadmaps for AI-related features in the next few quarters to anticipate potential risks.
- **What AI capabilities are being developed or integrated?**
  - Find out if AI is being built internally, procured from vendors, or embedded into existing systems.
- **Establish Quick Wins.**
  - While there may be no formalized program at the start, it'll be pivotal to establish quick wins for your team to show the importance of AI Risk Management.

Making Frameworks Work For You
(not the other way around)



- Most risk frameworks and guidance provided in the last few years have been vague to allow teams freedom of implementation
  - While this doesn't help with identifying the **exact** things you will need, you can start to to move your organization's maturity in the right direction
- **These are not one size fits all approaches**
  - Maybe you're an AI-native organization with a really good grasp on AI and now want to formalize governance.
  - Maybe you already have well-established or strong governance and are ready to build more technical demands of an AIRM program.

Defining Maturity with the MITRE

Security and Privacy

Protection of privacy rights and data security rights for AI is embedded and upheld by individuals designing, using, and overseeing AI systems to control the safety, specificity, and exchange of personal digital information.

Level 5	Enterprise-wide, AI-specific privacy and data security rights, approaches, and standards for individual control of safety, specificity, and exchange of digital data are improved and optimized based upon data trends.
Level 4	Enterprise-wide, AI-specific privacy and data security rights, approaches, and standards for individual control of safety, specificity, and exchange of digital information are managed and monitored.
Level 3	AI-specific privacy and data security rights, approaches, and standards defined and approved for individual control of safety, specificity, and exchange of digital information.
Level 2	AI-specific privacy and data security rights, approaches, and standards for individual control of safety, specificity, and exchange of digital information.
Level 1	No AI-specific privacy and data security rights, approaches, and standards are in place for individual security, control, safety, and specificity in exchange for digital information.

AI Innovation

Process whereby research, systems engineering, and human-machine design principles and best practices integrate AI into business operations.

Level 5	The approach to AI-specific systems engineering and Human-Machine practices is being refined to better achieve alignment with enterprise objectives.
Level 4	The approved approach to AI-specific systems engineering and Human-Machine practices is widespread and communicated, and metrics are being collected and analyzed to measure effectiveness.
Level 3	An approved approach to AI-specific systems engineering and Human-Machine practices is in place.
Level 2	An approach to AI-specific systems engineering and Human-Machine practices is being developed.
Level 1	No defined approach is present to research, identify, and utilize AI-specific systems engineering, and Human-Machine practices are applied in AI solutions.

Source: MITRE AI Maturity Model and Organizational assessment Tool

The NIST AI Playbook

- Where **MITRE** provided the language for cross functional alignment**, NIST** gives a high level overview of **functional program management.**

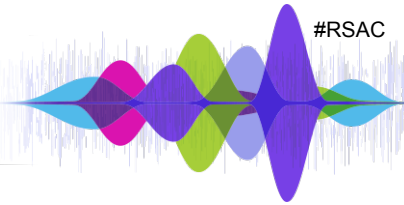
Map
Context is recognized and risks related to context are identified
Measure
Identified risks are assessed, analyzed, or tracked
Govern
A culture of risk management is cultivated and present
Manage
Risks are prioritized and acted upon based on a projected impact

Source: [NIST AI RMF Playbook]

Tailoring Existing Processes for AI-Driven Risk

- **Enhance existing security reviews and risk assessments** to integrate AI-specific concerns with minimal disruption.
- **Target high-impact** areas like application security and regulatory compliance to adapt proven processes without starting over.
- **Address AI-specific risks** swiftly while maintaining operational stability and supporting responsible innovation.
- **Strengthen data provenance and record keeping** to ensure traceability, support accountability, and enable more effective audits of AI systems.

Adapting Foundational Security Processes



You're already conducting security reviews, and you can extend these practices to AI

- **Automated Security Gates**
- **Security Champions**
- **Self-Service Security Review**
- **Ad-hoc Reviews**

Adapting Foundational Security Processes

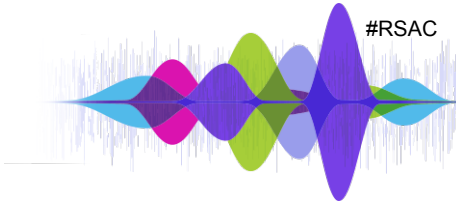
You're already reviewing your supply chain and third-party relationships, which can extend to inform your AIRM program

- **Open Source and SSDLC**
- **Usage and Licenses**
- **Dependency Tracking**

RSAC 2025 Conference

Looking Forward

Navigating a Changing Regulatory Landscape

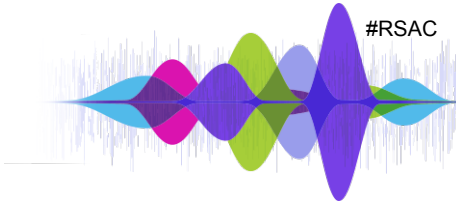


- **Engaging with your Legal partners early**
  - Provide an agile approach to adjust priorities as regulations shift.
  - Regularly assess finalized regulations to ensure compliance, even if it means playing catch up in some areas.
  - Continuous dialogue to integrate regulatory insights into your risk management strategy.
- **You can also do some homework yourself to stay informed**
  - Utilize publicly available resources and updates to stay informed about regulatory trends.
  - Stay involved in the community to see ongoing regulatory projects

Apply What You Have Learned Today

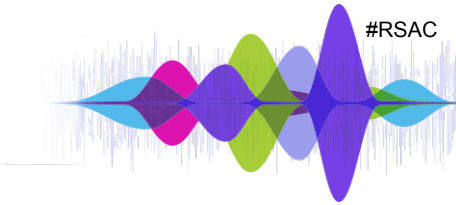
- **Next week you should:**
  - Run through the MITRE AI Maturity and do a light self-assessment
    - Establish and organizational baseline
    - Identify opportunities
    - Propose implementation

Apply What You Have Learned Today



- **In the first three months**
  - Find a few key AI projects in your organization that you can do a deep dive in partnership with Engineering teams
  - Document and analyze your findings to validate opportunities and refine your AI approach
    - ex. Identifying and streamlining lower-risk reviews vs. deep dives

Apply What You Have Learned Today



- **Within six months you should**:
  - Scale successful AI initiatives by expanding your deep dives to other high-impact areas
  - Establish a proactive process to monitor the regulatory environment
  - Engage cross-functional teams to standardize best practices

RSAC 2025 Conference

Many Voices.
One Community.

Questions?